ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2015) **13**(3); 195–205
doi:10.1017/S1479262114000884

Genetic diversity and population structure in a rice drought stress panel

Dindo A. Tabanao[1]*, Arnel E. Pocsedio[1], Jonalyn C. Yabes[1], Marjohn C. Niño[1], Reneth A. Millas[1], Neah Rosandra L. Sevilla[1], Xiao Yulong[2] and Jianming Yu[3]

[1]*Plant Breeding and Biotechnology Division, Philippine Rice Research Institute, Maligaya, Science City of Muñoz 3119, Nueva Ecija, Philippines,* [2]*Jiangxi Academy of Agricultural Science, No. 602 Nanlian Road, Nanchang, Jiangxi, P. R. China and* [3]*Department of Agronomy, Iowa State University, Ames, IA 50011, USA*

Received 21 May 2014; Accepted 13 August 2014 – First published online 12 September 2014

Abstract

A drought stress panel composed of diverse accessions selected from upland, aerobic, rainfed lowland and irrigated lowland environments, was assembled to serve as germplasm for aerobic adaptation breeding. Aerobic rice requires significant levels of tolerance to drought stress due to intermittent water deficit and high soil impedance caused by aerobic conditions. Genomic information may be utilized to investigate the nature of the panel to guide varietal improvement. Using 153 simple sequence repeat and 384 single nucleotide polymorphism markers, the aim of the study was to compare the allelic properties of the two marker types, infer population structure of the panel, and estimate kinship among the accessions. There was a general agreement between the results derived from the two marker types. Marker alleles were found to occur at low frequencies, as the panel was composed mostly of improved accessions with some landraces. The panel clustered into *japonica* (JA), *aus* (AU), upland-adapted *indica* (UL) and lowland-adapted *indica* (LL) subpopulations. The AU and JA subpopulations were more divergent from the rest of the subpopulations than were the LL and UL subpopulations. Average marker-based kinship for related accessions was less than 0.20, indicating a low degree of genetic relatedness in the panel. Within the LL and UL subpopulations, the low levels of kinship imply that there is still much genetic gain to be expected from utilizing the accessions in breeding. Thus, an understanding of the genetic variation in the panel suggests focusing on improving the mean in the short term, and tapping into the exotic alleles from the AU and JA subpopulations when genetic gain declines.

Keywords: genetic diversity; population structure; simple sequence repeats; single nucleotide polymorphisms; rice

Introduction

The success of a plant breeding programme depends on the availability of germplasm and the genetic variation that can be exploited from it. The use of modern cultivars in breeding for complex traits becomes limited due to their relatively narrow genetic base (Abdurakhmonov and Abdukarimov, 2008). In Asian rice (*Oryza sativa* L.), breeders can assemble a diverse breeding pool composed of landraces, improved lines and modern varieties that are spread across the main variety groups (*indica*, temperate *japonica*, tropical *japonica*, *aus* and aromatic; Garris *et al*., 2005, Zhao *et al*., 2011) to capture a wide range of adaptation and overcome the problem of low genetic base.

Rice is cultivated in a wide range of environments, such as irrigated lowland, rainfed lowland, upland,

* Corresponding author. E-mail: da.tabanao@philrice.gov.ph

flood-prone (Dixit *et al.*, 2012) and aerobic (Zhao *et al.*, 2010) rice areas. Irrigated lowland rice is grown in puddled soil throughout the cropping season; rainfed lowland rice relies entirely on rainfall or drainage from higher land areas; and, upland rice is grown in unbundled fields where soil drainage and land surface minimize the accumulation of water (Bernier *et al.*, 2008). Aerobic rice was more recently developed due to emerging global water scarcity threatening the sustainability of irrigated lowland rice production. Rice for aerobic culture requires high-yielding fertilizer-responsive genotypes incorporated with drought-tolerance traits to adapt to high soil impedance due to aerobic soil conditions (Lafitte and Bennett, 2002; Zhao *et al.*, 2010). Diverse breeding materials from these ecosystems were assembled to comprise the drought stress panel (DSP) utilized in this study to serve as germplasm for aerobic adaptation breeding.

The use of molecular markers in agricultural research has encouraged breeding institutions to adopt this technology as an integral part of their breeding programmes (Collard *et al.*, 2005). With the availability of abundant markers, genomic information allows a complementary strategy to using pedigree information for elucidating the nature of a breeding pool. In a diverse germplasm, data from genome-wide markers are useful in diversity analysis, population structure analysis and association studies (Agrama *et al.*, 2007; Huang *et al.*, 2010; Zhao *et al.*, 2011).

Two of the most promising marker systems, namely simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), have been widely used for genotyping in crops. SNPs have been known to occur in high density across the genome. They are amenable to high-throughput methods and have a simple mutation model. However, SNPs are biallelic and offer low information content (Frascaroli *et al.*, 2013). SSRs, on the other hand, have high allelic diversity and, in turn, have higher polymorphism rates than SNPs. The concurrent use of these two marker systems having different polymorphism mechanisms will provide results that are complementary to each other (Courtois *et al.*, 2012).

The objectives of the study were to (1) compare the allelic properties of SNP and SSR markers with respect to the DSP, (2) infer population structure of the breeding panel, and (3) estimate kinship among the accessions comprising the panel.

Materials and methods

Rice breeding panel and marker genotyping

The collection of 187 rice accessions used in this study (referred to as DSP) included 141 breeding lines and Philippine-released varieties, 27 local and foreign landraces, and 19 diversity checks. The breeding lines were entries from upland, aerobic, rainfed lowland and irrigated lowland nurseries of the International Network for Genetic Evaluation of Rice (INGER) of the International Rice Research Institute (IRRI). These materials were assembled to serve as breeding germplasm for genetic improvement of rice for adaptation to drought stress in aerobic culture.

Genomic DNA was extracted by the CTAB (cetyltrimethylammonium bromide) method (Murray and Thompson, 1980) from leaf tissue collected from single plants. For SNP genotyping, 20 μl containing at least 25 ng of DNA were prepared and submitted to the Genotyping Services Laboratory of IRRI in Laguna, Philippines. The RiceOPA3.1 (also known as *indica/japonica*) SNP set, a multiplex of 384 markers, was used (Thomson *et al.*, 2012). After excluding SNP markers with no polymorphism and more than 20% missing data, there remained 373 markers, with an average interval of 1.03 Mb between them.

For the SSR assay, polymerase chain reaction (PCR) was performed in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) thermocycler. The following temperature profile was used: initial denaturation at 95°C for 9 min, followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 52, 55 or 58°C for 2 min, and extension at 72°C for 2 min, and a final extension at 72°C for 7 min. Each PCR mix (7.5 μl) contained the following: 1.0 μl template DNA; 1.5 μl of 5 × Green GoTaq® Flexi Reaction Buffer (Promega Co., Madison, WI, USA); 0.25 μl of 25 mM $MgCl_2$; 0.4 μl of 5 mM dNTPs; 0.4 μl each of 10 μM forward and reverse primers; 0.2 μl (5 U/μl) of *Taq* recombinant DNA polymerase (Vivantis, Inc., Oceanside, CA, USA); 3.35 μl water. PCR products were fractionated in 4% (w/v) non-denaturing polyacrylamide gel and visualized by ethidium bromide staining. Bands were scored as codominant alleles. A total of 153 SSR markers were used for genotyping, with an average interval of about 2.5 Mb between markers.

Data analysis

Allele number, allele frequency and polymorphism information content (PIC), which is the proportion of the sum of squared allele frequencies at a marker locus (Anderson *et al.*, 1993), were calculated using PowerMarker 3.25 (Liu and Muse, 2005). At each SNP and SSR marker locus, whether there were two or more alleles, those that occurred less frequently than the most frequent allele were referred to as minor alleles.

The Bayesian model-based program STRUCTURE 2.3 (Falush *et al.*, 2003) was used to estimate the number

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

of subpopulations (k) given an admixture model with correlated allele frequencies. Simulations were run for 100,000 burn-in steps and 100,000 main run steps over ten independent iterations. The Delta k statistic of Evanno *et al.* (2005) was calculated using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). PowerMarker 3.25 (Liu and Muse, 2005) was used to calculate Nei's (1973) genetic distance and construct neighbour-joining trees (Saitou and Nei, 1987), which were visualized in MEGA5.2 (Tamura *et al.*, 2011).

Arlequin 3.5 (Excoffier *et al.*, 2005) was used to perform the analysis of molecular variance (AMOVA) and calculate pairwise genetic distance between subpopulations (Schneider and Excoffier, 1999) denoted as fixation index (F_{ST}). Statistical significance was tested by 1000 permutations. Rice accessions with an ancestry probability of less than 0.60 (according to model-based population structure analysis) were not included. The accessions that rearranged between subpopulations based on SNP data and those based on SSR data were also not included.

Marker-based genetic relatedness was calculated based on the Loiselle *et al.* (1995) formula for kinship (f) using SPAGeDi 1.3 (Hardy and Vekemans, 2002), given allele frequencies specific to the population. Negative estimates between any pair of genotypes, implying that the two were unrelated and with different ancestries (Thornton *et al.*, 2012), were set to 0. Estimates greater than 1 were set to 1. The minimum kinship coefficient to mean relatedness between any pair of accessions was set to 0.05, i.e. accessions with a kinship coefficient lower than 0.05 were considered unrelated.

Results

Allelic properties of SNP and SSR markers

For the RiceOPA3.1 SNP set, markers that detected only one allele were not included for data analysis, hence the average number of alleles was 2. For the 153 polymorphic SSRs, the number of alleles ranged from 2 to 7 with a mean of 3.47. Most of the markers (i.e. 127 out of 153, or 83%) amplified two to four alleles.

Most of the SNP and SSR alleles tended to occur at low rather than intermediate frequencies, although SSRs were slightly better enriched with intermediate allele frequencies. These low allele frequency levels resulted in frequency distributions that were skewed to the right (Fig. 1). The average minor allele frequency for SNPs was 0.125, compared with 0.148 for SSRs, which was only slightly higher.

In terms of PIC, the frequency distributions for both SNP and SSR markers were less skewed than that of allele frequency. PIC approximately centred around the mean of 0.17 for SNP markers and around the mean of 0.43 for SSR markers.

Population structure

Mean values of L(k) (i.e. natural logarithm of the posterior probability of k given the data) for $k = 1–10$, where k was the given number of subpopulations, showed that the increase in L(k) slowed sharply above $k = 2$, which was also reflected in the very high peak at $k = 2$ in the Delta k plots. This subdivision corresponds to the very distinct separation between *indica* and *japonica* (JA) accessions (Fig. 2). At $k = 3$, the *aus* (AU) accessions (representing *aus*-type rice) differentiated from the rest of the *indica* accessions for SNPs, whereas a group of upland rice (UL) accessions differentiated from the rest of the *indica* accessions for SSRs. At $k = 4$, the UL accessions differentiated from the remaining *indica* accessions (which were mostly irrigated lowland and rainfed lowland accessions) for SNPs, whereas the AU group diverged from the remaining *indica* accessions for SSRs. At $k = 5$, the remaining *indica* accessions diverged into two groups for both markers, but there was no apparent pattern to this subdivision, as both groups comprised a mixture of mostly rainfed lowland and irrigated lowland accessions. For this reason, the number of biologically plausible subpopulations was set to 4. Because the rainfed lowland and irrigated lowland rice accessions could not be clearly differentiated from each other, the lowland-adapted *indica* (LL) was a disproportionately large subpopulation, comprising at least 53% of the DSP.

Ancestry probability was less than 0.60 for 32 accessions (16 based on SNPs and 21 based on SSRs), which were categorized as having admixed ancestry. Most of these accessions were an admixture of LL and UL ancestries. These were breeding lines developed for upland environments, which were often crosses between tropical *japonica* genotypes (to impart upland adaptation) and irrigated lowland-adapted lines (to impart high yield potential).

Only one accession was assigned to two different subpopulations depending on the marker type. TCA80-4, a breeding line from Africa, had the following ancestry probabilities: 0.66 LL:0.34 UL based on SNP marker data and 0.39 LL:0.60 UL based on SSR marker data. Apparently, this genotype was a combination of two different genetic backgrounds, but obtained an ancestry probability of at least 0.60 from either marker type such that it was not classified as having an admixed ancestry.

The agreement in subpopulation assignment between SSR and SNP markers was remarkably high. Of the 32 accessions with admixed coancestry between the SNP and SSR markers, five were classified as such by both

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

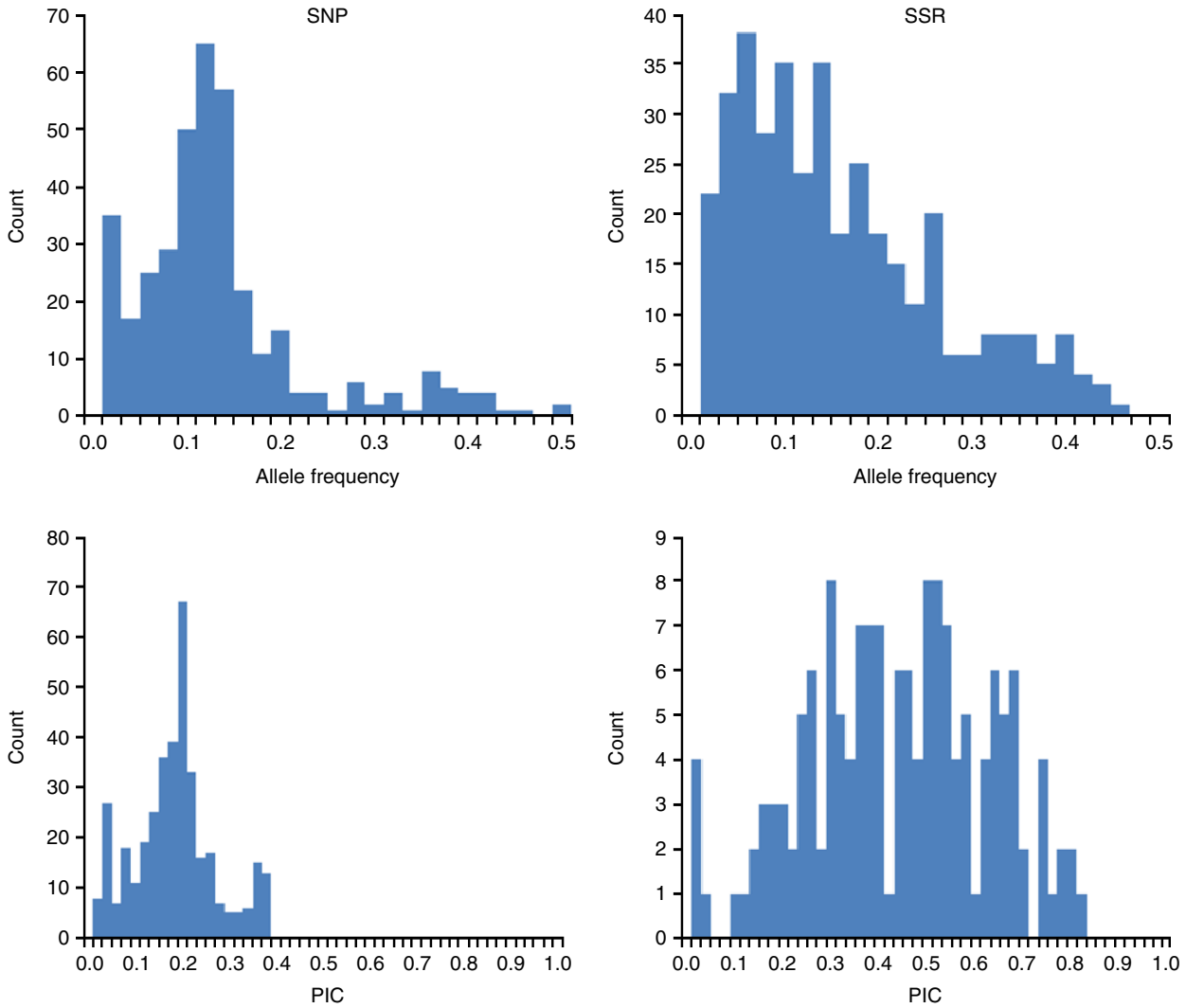

Fig. 1. Frequency distribution of minor allele frequency and PIC for SNP and SSR markers.

marker types. From the remaining 155 accessions with an ancestry probability of at least 0.60, only one was rearranged between the two marker types.

Cluster analysis

Except for the difference in resolution in the *indica* main cluster, the branching patterns between the SNP- and SSR-based neighbour-joining trees were in general agreement with each other (Fig. 3). In both SNP- and SSR-based trees, the *indica* group (UL and LL) clustered in one direction, whereas the JA (red) accessions were in the opposite direction. It can also be seen in both trees that the UL accessions (orange) separated from the rest of the *indica* (blue). The AU accessions formed a separate cluster in both trees. In the SNP-based tree, the AU cluster was in the same direction with the *indica* group, while in the SSR-based tree, it was midway between the *indica* and JA main branches. Both trees also depicted that there was no apparent structure within the large LL subpopulation.

In general, the major clusters of the neighbour-joining trees (Fig. 3) agreed with the subpopulations based on model-based groups. Some of the LL accessions were interspersed in some of the predominantly UL branches, and vice versa. Admixed ancestry rice accessions were mostly interspersed with the UL accessions, and some may be found along the juncture between the *indica* and JA subdivision, and among the LL accessions.

Genetic differentiation among subpopulations

The eight accessions that were rearranged between the subpopulations based on SNP and SSR data were not included in the population differentiation analysis.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

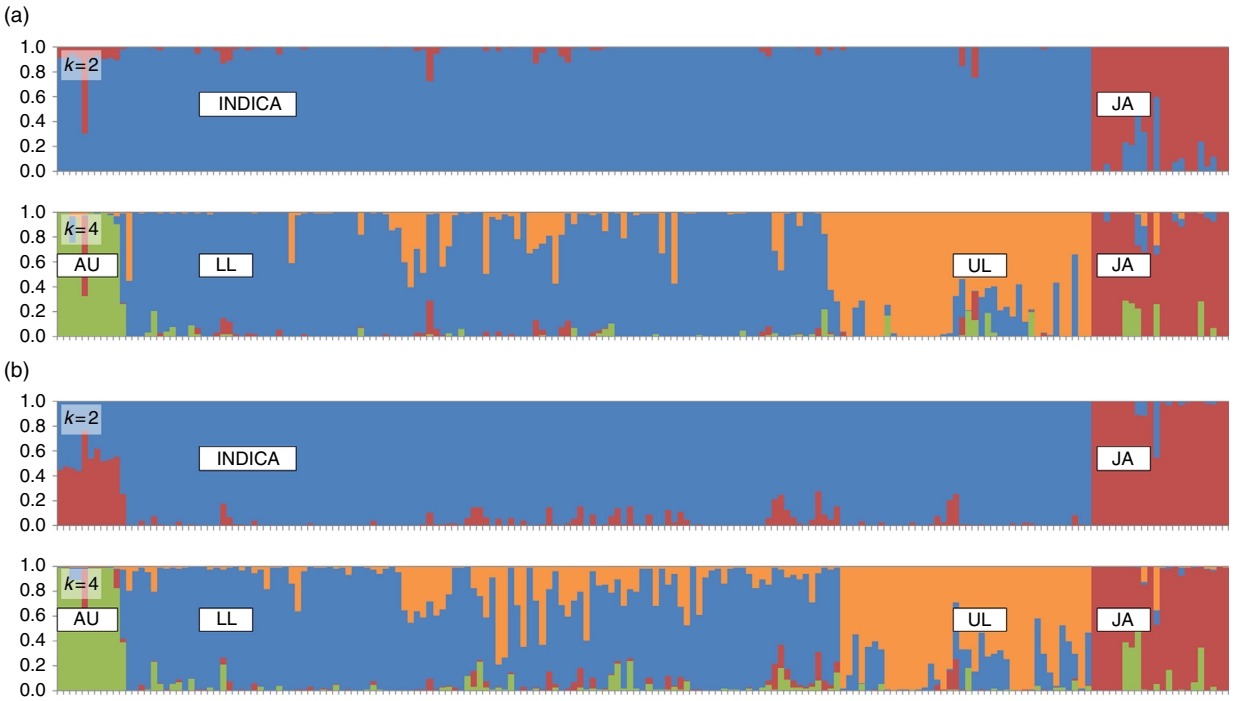

Fig. 2. Ancestry probability of 187 rice accessions inferred by STRUCTURE using (a) 373 SNP and (b) 153 SSR markers at $k = 2$ and $k = 4$ subpopulations.

The 32 accessions with admixed ancestry (16 based on SNP data and 21 based on SSR data) were also not included. In total, there were 33 accessions that were either rearranged or had admixed ancestry, leaving 154 accessions across four subpopulations (AU, LL, UL and JA) that were analysed for population differentiation.

The results of the AMOVA of the SNP data showed that most of the molecular diversity in the panel was due to variation among subpopulations (75.9%), and only about one-quarter was due to the variation within the subpopulations (Table 1). The overall F_{ST} among the model-based groups was 0.7588, implying a relatively high level of differentiation. Except for the LL and UL subpopulations, which had a relatively low F_{ST} (0.2773), the subpopulations were highly differentiated from each other, with pairwise F_{ST} ranging from 0.7097 to 0.8580 (Table 2).

The overall F_{ST} based on SSR data was much lower ($F_{ST} = 0.2920$), and the results of AMOVA showed a reversed trend: molecular diversity due to the variation among the subpopulations (29.2%) was lower than the within-subpopulation variation (70.8%). In other words, according to the information obtained by the SSR assay, the main *indica* and JA groups were individually more diverse than the SNP assay would suggest.

The levels of genetic differentiation among the model-based subpopulations were lower for SSRs than for SNPs. Pairwise F_{ST} ranged from 0.1206 to 0.4429. The lowest F_{ST} was observed between the LL and UL subpopulations, as what was observed in the SNP-based F_{ST}. The lowest level of differentiation being observed between the LL and UL subpopulations (for both SNP and SSR markers) was apparently because most of the accessions belonging to these subpopulations were interrelated, as they were breeding lines developed from inter-crosses between the irrigated lowland, rainfed lowland and upland-adapted genotypes.

Kinship

Marker-based kinship (f) estimates in the DSP were mostly less than 0.05, and pairs of accessions with such kinship coefficient were considered unrelated. For related accessions, the mean kinship was 0.160 based on SNP data, and 0.187 based on SSR data. In other words, kinship estimates for the most part were very close to 0 (Fig. 4, top), implying that the degree of genetic relatedness in the DSP was generally low.

The relatedness between rice accessions was apparent within the subpopulations, especially in the AU and JA subpopulations where mean kinship was high (0.598 based on SNP data and 0.469 based on SSR data). This high level of kinship was also depicted by the deep red boxes on the heat map corresponding to the AU and JA subpopulations (Fig. 4, bottom). However, these two subpopulations represented only a small proportion (17%) of the whole panel.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

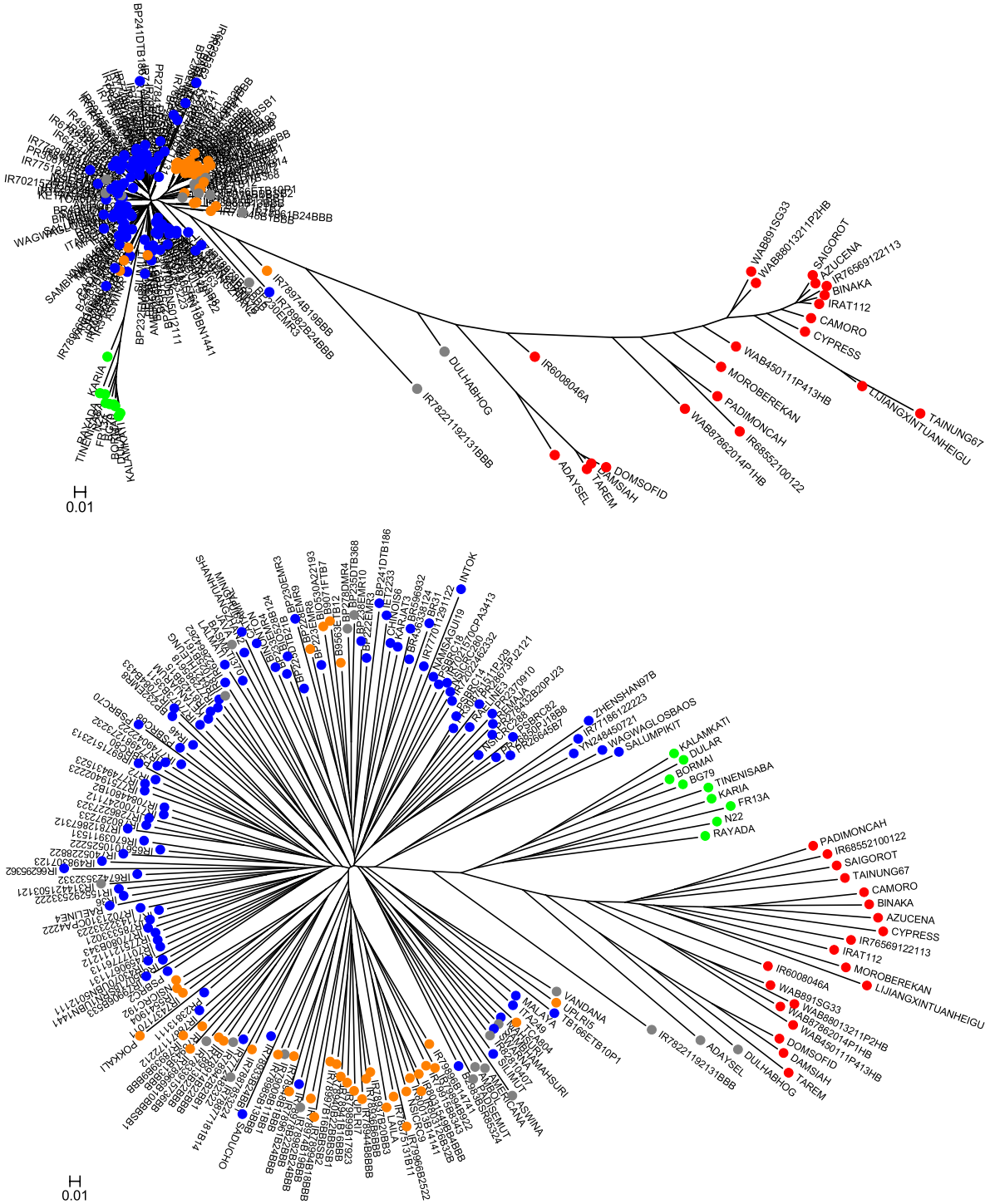

Fig. 3. Neighbour-joining trees of 187 rice accessions based on Nei's (1973) genetic distance using SNP (top) and SSR marker (bottom) data. The colour of each accession corresponds to subpopulation assignment by STRUCTURE (AU, green; LL, blue; UL, orange; JA, red; admixed ancestry, grey).

There was some degree of inter-relatedness between the members of the LL and UL subpopulations, and between those of the AU and JA subpopulations, as indicated by the faint red boxes between the subpopulations in Fig. 4. The genetic relatedness between the members of the LL and UL subpopulations corroborated the relatively lower degree of genetic differentiation between these two subpopulations, as shown in Table 2.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

Table 1. Analysis of molecular variance of SNP and SSR marker data

	SNP		SSR	
Source of variation	Variance components	Percentage of variation	Variance components	Percentage of variation
Among subpopulations	43.41	75.88	11.22	29.20
Within subpopulations	13.80	24.12	27.20	70.80
Total	57.21		38.42	

Pearson's correlation between pairwise kinship based on SNP and SSR markers was moderately high ($r = 0.78$, $P < 0.001$), implying a general agreement between, and the suitability of the Loiselle *et al.* (1995) formula to, the two marker types. Yu *et al.* (2009) preferred this formula because it does not assume Hardy–Weinberg equilibrium, a prerequisite that cannot be assumed in the DSP as well. In Fig. 4, a similar pattern could be observed in the heat maps of kinship based on SNP and SSR markers. Studies from other crops also showed general agreement between SNP- and SSR-based genetic diversity estimates (Van Inghelandt *et al.*, 2010; Yang *et al.*, 2011; Würschum *et al.*, 2013).

Discussion

Genetic diversity and population structure

The use of two independent marker systems, namely SSRs and SNPs, successfully characterized the genetic diversity and population structure of the DSP. The allele frequencies in the panel were found to be low in both marker types. In maize, allele frequencies of SNPs were reported to be distributed evenly from 0 to 0.5 in improved lines and cultivars (Hamblin *et al.*, 2007; Yang *et al.*, 2011). The same observation was reported for the self-pollinated crop wheat (Würschum *et al.*, 2013). In the present study, the population contained some genotypes that were either landraces or improved non-*indica* accessions, which may carry alleles throughout the genome that were not in common with the rest of the accessions. The outcome would be an excess of low-frequency alleles even for biallelic SNP markers. Consequently, PIC values were generally lower than what have been previously reported because of the composition of the panel, being mostly improved lines and cultivars with some landraces. For SSR markers, PIC would tend to be not lower than 0.50 when the sample of genotypes is composed mostly of landraces (Garris *et al.*, 2005; Ram *et al.*, 2007; Pervaiz *et al.*, 2009). PIC takes into account not only the number of alleles, but also their frequencies. If many alleles are present only in low frequencies, as was the case in the DSP, the PIC is not expected to be high.

In general, the major clusters of neighbour-joining trees agreed with the subpopulations based on model-based grouping. The breeding panel was structured into JA, AU, UL and LL. When global diversity is considered, the number of known subpopulations is 5, representing the five main variety groups of Asian rice: *indica*; tropical *japonica*; temperate *japonica*; AU; aromatic (Garris *et al.*, 2005; Zhao *et al.*, 2011). In contrast, the DSP did not represent global diversity, hence the observed clustering did not correspond to the five known variety groups.

The results of AMOVA showed a reversed trend between the SNP and SSR markers. SNP data showed that most of the molecular diversity in the panel was due to the variation among the subpopulations, while SSR data showed that the within-subpopulation variation was higher than the among-subpopulation variation. One explanation is the way the RiceOPA3.1 SNP set was designed, which was to differentiate *indica* from JA genotypes, such that the emphasis was variation between, and not within, the two groups. Another reason is the difference in the basis of polymorphism between the two marker types. Single-base changes occur much more slowly through time, reflecting deeper variation between genotypes. In contrast, SSRs, which are due to polymerase slippage, have a higher mutation rate (Hamblin *et al.*, 2007) and therefore reflect more recent variation in DNA.

Kinship estimates were generally low. Genetic relatedness estimates between accessions were higher within, than across, the subpopulations. There was some degree of inter-relatedness between the members of the

Table 2. Pairwise F_{ST} based on SNP (below diagonal) and SSR markers (above diagonal) among the four subpopulations[a]

	AU	LL	UL	JA
AU		0.2833	0.3636	0.3744
LL	0.7097		0.1206	0.3967
UL	0.7348	0.2773		0.4429
JA	0.7820	0.8580	0.8186	

[a] All estimates were significant ($P < 0.001$) based on 1000 permutations.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

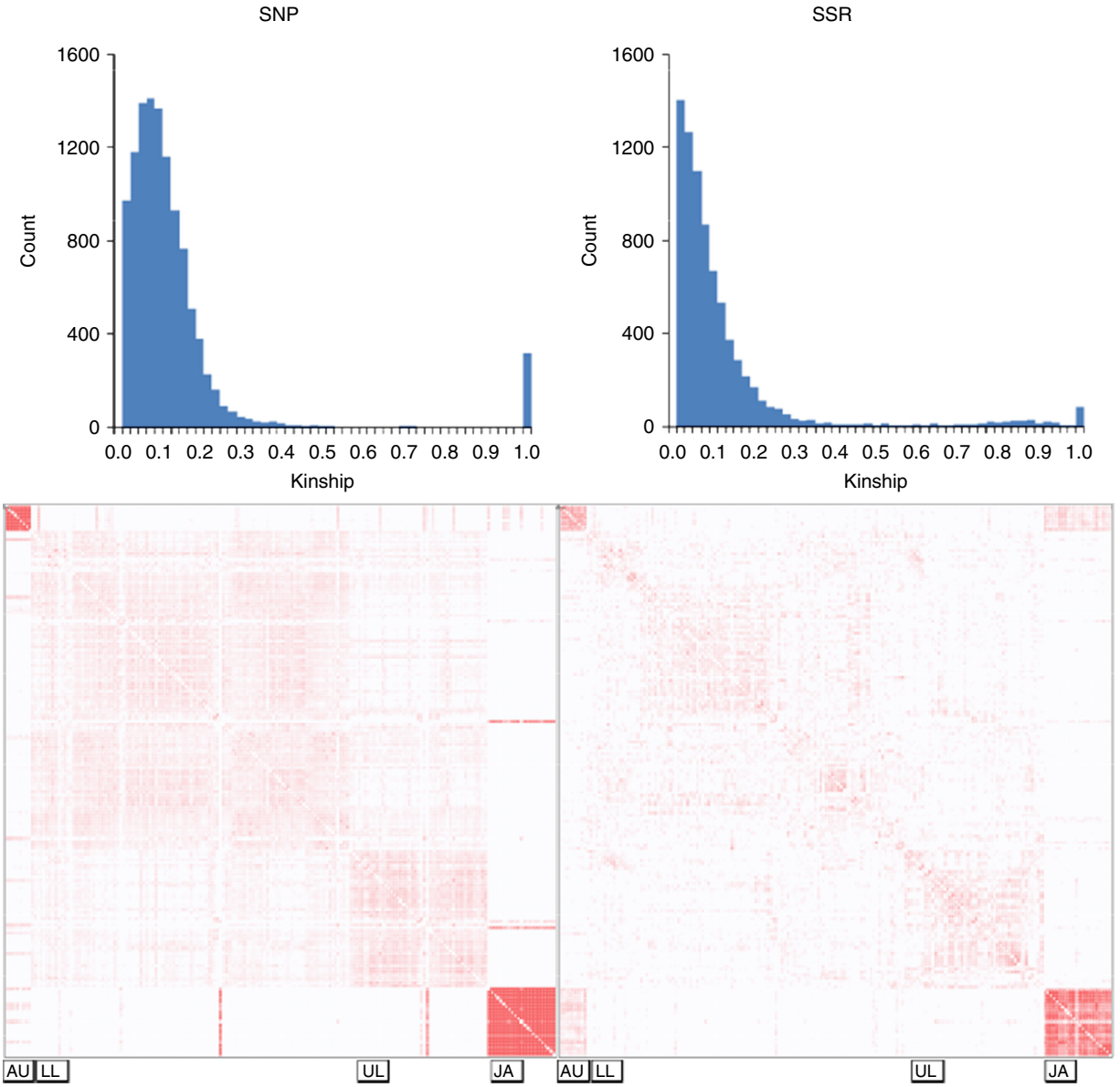

Fig. 4. Kinship (*f*) based on the Loiselle *et al.* (1995) formula. Top: frequency distributions of kinship estimates among related (i.e. $f \geq 0.05$) accessions based on the SNP (left) and SSR (right) markers. Bottom: kinship matrices based on the SNP (left) and SSR (right) markers. Accessions are arranged according to model-based subpopulation assignment.

LL and UL subpopulations, which corroborated the relatively lower degree of genetic differentiation between these two subpopulations. The lack of inter-relatedness among the accessions from different subpopulations (such as between the LL and JA groups) was also observed. These observations highlighted the difference between kinship and population structure. Familial relatedness is significant only in recent generations, whereas population structure is due to ancestry differences that have developed through long evolutionary time (Yu *et al.*, 2005). Genetic relatedness within the LL and UL subpopulations was observed to be low, implying that these breeding lines were genetically diverse, and that there is much opportunity for genetic improvement. The development of advanced cycle inbreds by crossing elite by elite lines in pedigree breeding is one of the reasons for genetic gain to decline over time. As for the LL and UL accessions in this breeding panel, loss of variation is not a problem as yet. Thus, the immediate breeding strategy is to focus on improving the mean by crossing the improved *indica* lines within the LL or UL subpopulations, and to tap into the exotic alleles of the AU and JA subpopulations later, when genetic gain declines.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press

SSR markers, being often able to detect more than two alleles, provide more information than do SNP markers. Hamblin *et al.* (2007) compared 847 SNP with 89 SSR markers in terms of assigning maize inbreds to subpopulations, and concluded that the small number of SSR markers exhibited higher discriminatory power over the biallelic SNP markers. Van Inghelandt *et al.* (2010) concluded that there should be seven to ten times more SNP markers to achieve the precision of SSR markers in the analysis of genetic diversity and population structure in maize. In the present study, the RiceOPA3.1 SNP set seemed to have provided the same degree of efficiency as the 153 SSRs in assigning traditional and improved accessions into different subpopulations and in determining their genetic relatedness. The two marker types complemented each other in assessing the variation between and within subpopulations (i.e. genetic differentiation between subpopulations).

Upland rice in improved genetic background

The results from the use of both SNP and SSR markers pointed to the formation of the UL subpopulation, which was distinct from, but which seemed to be an intermediate group between, the LL and JA subpopulations. Traditional rice varieties that are adapted to upland culture are mostly tropical *japonica* (Garris *et al.*, 2005; Atlin *et al.*, 2006); however, UL accessions in this study were found to be closer to the *indica* group than to the JA group. In fact, in the Bayesian model-based clustering, UL accessions belonged to the *indica* group, not to the JA group, when the number of subpopulations was set to 2. Whereas UL accessions have the adaptation to upland environments as JA accessions generally do, their yielding ability and plant stature are more similar to those of improved accessions belonging to LL. In other words, UL accessions are upland-adapted genotypes in a modern genetic background as a result of varietal improvement.

Zhao *et al.* (2010) reported a new production system that grows rice in non-puddled aerobic soils (i.e. upland environment), but which utilizes a different set of varieties called aerobic rice. Such genotypes are a product of breeding that utilizes traditional upland varieties, to confer drought tolerance, and modern irrigated lowland-adapted rice varieties, which are high yielding and input-responsive. This has been the strategy employed by breeding programmes to improve the productivity of uplands. Bernier *et al.* (2007) and Venuprasad *et al.* (2012) also reported aerobic breeding efforts that concentrated on identifying quantitative trait loci from upland-adapted entries in semi-dwarf, high-yielding *indica* backgrounds. In this study, the 32 accessions that were considered admixed (i.e. ancestry probability <0.60) became as such because they were also mostly an admixture of LL and UL ancestries. Indeed, based on cluster analysis, they tended to intersperse with UL accessions, as shown in Fig. 3. The separation of UL accessions pointed to the formation of a new group and the admixed accessions could represent breeding materials that were transitioning to UL. Previous studies (Garris *et al.*, 2005; Lu *et al.*, 2005; Giarocco *et al.*, 2007) attributed genetic structure in different panels of rice accessions to domestication events. In this study, genetic structure, particularly for UL accessions, was found to be a result of modern breeding. Courtois *et al.* (2012) and Shinada *et al.* (2014) similarly found that plant breeding programmes have led to the formation of new genotypes with a unique morphology and ecosystem-specific adaptation. The analysis of genome-wide genotype data affords the profound understanding of genetic variation to guide the conceptualization and realization of breeding strategies for adaptation to less favourable environments, such as drought-prone and aerobic rice areas.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1479262114000884

Acknowledgements

This study was funded by the Philippine Government through the Philippine Rice Research Institute and the Department of Agriculture National Rice Program.

References

Abdurakhmonov IY and Abdukarimov A (2008) Application of association mapping to understand the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics* 2008: 1–18.

Agrama HA, Eizenga GC and Yan W (2007) Association mapping of yield and its components in rice cultivars. *Molecular Breeding* 19: 341–356.

Anderson JA, Churchill GA, Autrique JE, Tanksley SD and Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36: 181–186.

Atlin GN, Lafitte HR, Tao D, Laza M, Amante M and Courtois B (2006) Developing rice cultivars for high-fertility upland systems in the Asian tropics. *Field Crops Research* 97: 43–52.

Bernier J, Kumar A, Venuprasad R, Spaner D and Atlin G (2007) A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. *Crop Science* 47: 507–518.

Bernier J, Atlin G, Serraj R, Kumar A and Spaner D (2008) Breeding upland rice for drought resistance. *Journal of the Science of Food and Agriculture* 88: 927–939.

Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142: 169–196.

Courtois B, Frouin J, Greco R, Bruschi G, Droc G, Hamelin C, Ruiz M, Clément G, Evrard JC, Coppenole S, Katsantonis D, Oliveira M, Negrão S, Matos C, Cavigiolo S, Lupotto E, Piffanelli P and Ahmadi N (2012) Genetic diversity and population structure in a European Collection of Rice. *Crop Science* 52: 1663–1675.

Dixit S, Swamy BPM, Vikram P, Ahmed HU, Sta Cruz MT, Amante M, Atri D, Leung H and Kumar A (2012) Fine mapping of QTLs for rice grain yield under drought reveals sub-QTLs conferring a response to variable drought severities. *Theoretical and Applied Genetics* 125: 155–169.

Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.

Excoffier L, Laval G and Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1: 47–50.

Falush D, Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.

Frascaroli E, Schrag TA and Melchinger AE (2013) Genetic diversity analysis of elite Europen maize (*Zea mays* L.) inbred lines using AFLP, SSR and SNP markers reveals ascertainment bias for a subset of SNPs. *Theoretical and Applied Genetics* 126: 133–141.

Garris AJ, Tai TH, Coburn J, Kresovich S and McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169: 1631–1638.

Giarocco LE, Marassi MA and Salerno GL (2007) Assessment of the genetic diversity in Argentine rice cultivars with SSR markers. *Crop Science* 47: 853–860.

Hamblin MT, Warburton ML and Buckler ED (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS ONE* 2(12): e1367.

Hardy OJ and Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li Jiayang and Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42: 961–969.

Lafitte HR and Bennett J (2002) Requirements for aerobic rice: physiological and molecular consideration. In: Bouman BAM, Hengsdijk H, Hardy B, Bindraban PS, Tuong TP and Ladha JK (eds) *Proceedings of the International Workshop on Water-Wise Rice Production*, 8–11 April 2002. Los Baños, Philippines: International Rice Research Institute, pp. 259–274.

Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.

Loiselle BA, Sork VL, Nason J and Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* 82: 1420–1425.

Lu H, Redus MA, Coburn JR, Rutger JN, McCouch SR and Tai TH (2005) Population structure and breeding patterns of 145 U.S. rice cultivars based on SSR marker analysis. *Crop Science* 45: 66–76.

Murray MG and Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4325.

Nei M (1973) The theory and estimation of genetic distance. In: Morton ME (ed.) *Genetic Structure of Populations*. Honolulu: University Press of Hawaii, pp. 45–54.

Pervaiz ZH, Rabbani MA, Pearce SR and Malik SA (2009) Determination of genetic variability of Asian rice (*Oryza sativa* L.) varieties using microsatellite markers. *African Journal of Biotechnology* 8: 5641–5651.

Ram SG, Thiruvengadam V and Vinod KK (2007) Genetic diversity among cultivars, landraces and wild relatives of rice as revealed by microsatellite markers. *Journal of Applied Genetics* 48: 337–345.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.

Schneider S and Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 152: 1079–1089.

Shinada H, Yamamoto T, Yamamoto E, Hori K, Yonemaru J, Matsuba S and Fujino K (2014) Historical changes in population structure during rice breeding programs in the northern limits of rice cultivation. *Theoretical and Applied Genetics* 127: 995–1004.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung CW, Reynolds A, Scheffler B, Eizenga G, McCLung A, Kim H, Ismail AM, de Ocampo M, Mojica C, Reveche MY, Dilla-Ermita CJ, Mauleon R, Leung H, Bustamante C and McCouch SR (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding* 29: 875–886.

Thornton T, Tang H, Hoffman TJ, Ochs-Balcom HM, Caan BJ and Risch N (2012) Estimating kinship in admixed populations. *American Journal of Human Genetics* 91: 122–138.

Van Inghelandt D, Melchinger AE, Lebreton C and Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theoretical and Applied Genetics* 120: 1289–1299.

Venuprasad R, Bool ME, Quiatchon L and Atlin GN (2012) A QTL for rice grain yield in aerobic environments with large effects in three genetic backgrounds. *Theoretical and Applied Genetics* 124: 323–332.

Würschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E and Reif JC (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theoretical and Applied Genetics* 126: 1477–1486.

Yang X, Xu Y, Shah T, Li H, Han Z, Li J and Yan J (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetics* 139: 1045–1054.

Yu J, Pressoir G, Briggs WH, Bi Y IV, amasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S and Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203–208.

Yu J, Zhang Z, Zhu C, Tabanao DA, Pressoir G, Tuinstra MR, Kresovich S, Todhunter RJ and Buckler ES (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. *Plant Genome* 2: 63–77.

Zhao DL, Atlin GN, Amante M, Sta Cruz MT and Kumar A (2010) Developing aerobic rice cultivars for water-short irrigated and drought-prone rainfed areas in the tropics. *Crop Science* 50: 2268–2276.

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD and McCouch SR (2011) Genome-wide association mapping reveals a rich genetic structure of complex traits in *Oryza sativa*. *Nature Communication* 2: 467.

https://doi.org/10.1017/S1479262114000884 Published online by Cambridge University Press